

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problems Mailbox.**

JC997 U.S. PRO  
09/942615



For: Our New U.S. Patent Application. Our Ref: NT0451US

[Media Distribution Systems and Multi-media conversion server

(J. Kimura; et al)

LIST OF THE PRIOR ART REFERENCES CITED IN THE SPECIFICATION.

1. Japanese Laid-Open No.06-162167 (COMPOSITE IMAGE DISPLAY SYSTEM), (Date of publication : 10.06.1994) with English abstract thereof.
2. Japanese Laid-Open No.07-066832 ( Multimedia Electronic Mail System), (Date of publication : 10.03.1995) with English abstract thereof.

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 06-162167

(43)Date of publication of application : 10.06.1994

(51)Int.Cl.

G06F 15/62  
G06F 3/153  
G06F 3/16  
G10L 3/00

(21)Application number : 04-335526

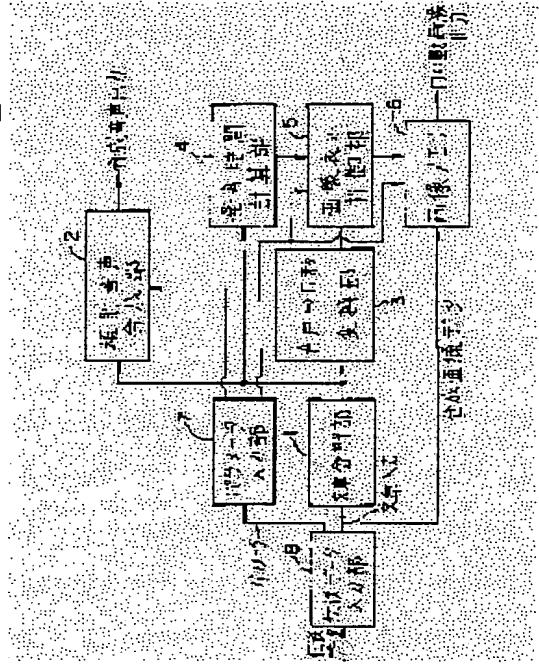
(71)Applicant : FUJITSU LTD

(22)Date of filing : 20.11.1992

(72)Inventor : NAKAGAWA AKIRA  
MORIMATSU EIJI  
MATSDA KIICHI**(54) COMPOSITE IMAGE DISPLAY SYSTEM****(57)Abstract:**

**PURPOSE:** To make a display exactly as a sending-side person intends by embedding parameters of a composite image that a person who composes data desires on an expressing side in the composite data on a face image and using the values embedded in the image data as initial values of a system on a display side.

**CONSTITUTION:** Mapping to a face model is performed on the basis of the original image of the desired face image to be displayed on a reception side, and parameters of respective mouth shapes are used to generate the composite image data, in which the parameters regarding an impression to be given to the opposite reception side are embedded. Document information is generated separately and the both are sent as transmitted data to the reception side. The data are inputted to a transmitted data input part 8; and document information is sent to a document decomposition part 1, the composite image data is sent to an image memory 6, and various parameters are sent to a parameter input part 7. The parameter input part 7 once receiving the parameters checks and sends them to an image display control part 5 and the image memory 6. On the reception side, the information is displayed exactly as the sending-side person intends.

**LEGAL STATUS**

[Date of request for examination] 19.11.1999

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平6-162167

(43)公開日 平成6年(1994)6月10日

(51)Int.Cl. <sup>5</sup>	識別記号	府内整理番号	F I	技術表示箇所
G 0 6 F 15/62	3 4 0	8125-5L		
3/153	3 2 0	L 7165-5B		
3/16	3 3 0	C 7165-5B		
G 1 0 L 3/00	S	8946-5H		

審査請求 未請求 請求項の数3(全7頁)

(21)出願番号 特願平4-335526

(22)出願日 平成4年(1992)11月20日

(71)出願人 000005223  
富士通株式会社  
神奈川県川崎市中原区上小田中1015番地  
(72)発明者 中川 章  
神奈川県川崎市中原区上小田中1015番地  
富士通株式会社内  
(72)発明者 森松 映史  
神奈川県川崎市中原区上小田中1015番地  
富士通株式会社内  
(72)発明者 松田 喜一  
神奈川県川崎市中原区上小田中1015番地  
富士通株式会社内  
(74)代理人 弁理士 小林 隆夫

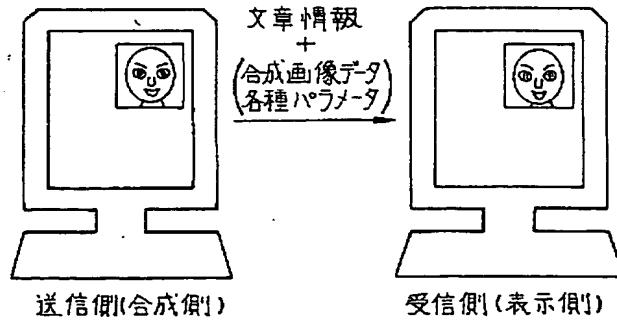
(54)【発明の名称】 合成画像表示システム

(57)【要約】

【目的】 文章(テキスト)データを送るだけであったかもTV電話のように送信者が喋っている顔の合成動画像と合成音声で相手側にメッセージを伝えることができるAV(オーディオ・ビデオ)電子メール等に適用できる合成画像表示システムに係り、特に、文章作成側で意図した声質や顔表情等の印象を的確に表示側に反映させることができる合成画像表示システムに関するものであり、表示側で文章情報に基づいて合成音声あるいは顔合成画像を表示するにあたり、その文章の作成側の人が意図した通りの表示が可能となるようにすることを目的とする。

【構成】 任意の文章情報からそれに対応する合成音声および合成音声に合わせて口が動く人物の顔の合成動画像を生成する合成画像表示システムにおいて、文章情報の作成側において顔の合成画像を作成する際に、表示側における合成音声と合成動画像の生成態様を決めるための各種パラメータを合成画像データに付加して表示側に渡すように構成されたことを特徴とする。

本発明に係る原理説明図



**【特許請求の範囲】**

**【請求項1】** 任意の文章情報からそれに対応する合成音声および該合成音声に合わせて口が動く人物の顔の合成動画像を生成する合成画像表示システムにおいて、文章情報の作成側において顔の合成画像を作成する際に、表示側における合成音声と合成動画像の生成態様を決めるための各種パラメータを該合成画像に付加して表示側に渡すように構成されたことを特徴とする合成画像表示システム。

**【請求項2】** 該各種パラメータは合成音声の声質、合成動画像を表示する際の表示倍率、表示位置を含むパラメータである請求項1記載の合成画像表示システム。

**【請求項3】** 受信した伝送データを合成動画像データ、文章情報、各種パラメータに分離する伝送データ入力手段と、

該文章情報に基づいて合成音声を生成し出力する音声合成手段と、

該伝送データ入力手段で分離された合成動画像データをファイリングする画像メモリと、

該文章情報をその文章情報を発声したときの一連の口形の動きを表す口形符号の系列に変換する変換手段と、該文章情報に基づいて該音声合成手段から出力される合成音声の各音節の発音時間を計算して各音声の切れ目のタイミングを推定する発音時間計算手段と、

該発音時間計算手段で推定した各音節の切れ目のタイミングで表示画像を該変換手段からの口形符号に対応した口形画像に切り換える制御を行う表示制御手段と、該伝送データ入力手段で分離された各種パラメータを対応する内部回路に送るパラメータ入力手段とを備えた合成画像表示システム。

**【発明の詳細な説明】**

**【0001】**

**【産業上の利用分野】** 本発明は、文章（テキスト）データを送るだけであったかもTV電話のように送信者が喋っている顔の合成動画像と合成音声で相手側にメッセージを伝えることができるAV（オーディオ・ビデオ）電子メール等に適用できる合成画像表示システムに係り、特に、文章等の作成側で意図した声質や顔表情等の印象を表示側に伝えることができる合成画像表示システムに関するものである。

**【0002】**

**【従来の技術】** 任意の文章（テキスト）情報からそれに応じた合成音声を自由に生成し発音する技術は、規則音声合成と呼ばれ、これを実現するための規則音声合成装置が既に作られている。この規則音声合成技術は人間と機械とのインターフェースを向上させるために様々な分野で応用されている。また、近年、音声の合成と同様に、任意の文章情報からそれを喋ったときの口の動きを含む人物の動画像をその文章情報を解析することで生成する技術が開発されており、これを上述の音声合成技術

と組み合わせることによって、より自然なインターフェースを実現することができる。

**【0003】** 例えば、かかる音声と顔動画像の合成技術を電子メールに適用すると、受信側にメール送信者の顔画像などのデータファイルを予め用意しておくことにより、従来では受信側の画面上に文章が表示されるだけであった電子メールに対して、メール送信者が喋っている顔の動画像が現れて合成音声で読み上げるといった表現豊かなメッセージを受信者に伝えることができる。

**【0004】** このような文章に基づいて音声および顔動画像を合成し出力する音声・動画像出力装置の構成例を図4に示す。図4において、1は文章（テキスト）情報が入力される文章分解部であり、この文章分解部1は入力された文章情報を解析して音声出力用の発音制御データを生成し規則音声合成部2と音声／口形変換部3に出力する。例えば、文章情報として「ただいま」の文章が入力された場合、これを「T, A, D, A, I, M, A」の母音と子音からなる音素データに分解して出力する。

**【0005】** 規則音声合成部2は任意の文章についての音素データに基づいてその文章を読み上げる合成音声を生成し出力する装置である。

**【0006】** 音声／口形変換部3は、任意の文章についての音素データをその文章を発音する際の一連の口の動きを表すための口形符号の系列に変換するための装置である。口形符号としては例えば、A（母音のア）、I（母音のイ）、U（母音のウ）、E（母音のエ）、O（母音のオ）、S（子音）、C（閉じた口）の7種類があり、それぞれの口形符号に対応してそれらを発音する際の口形の画像が予め用意される。例えば、文章情報として前述の「ただいま」の文章が入力された場合、その文章の音素データ「T A D A I M A」に基づいて、「T」→口形符号S、「A」→口形符号A、「D」→口形符号S、「I」→口形符号I、「M」→口形符号C、「A」→口形符号A、をそれぞれ割り当てて、それらを口形符号の系列として画像表示制御部5に出力する。

**【0007】** 画像メモリ6には合成画像データがファイリングされている。この合成画像データとしては、話者の1フレーム分の肩上画像と、それを基に合成した前述の7種類の口形符号に対応した7種類の口領域画像のデータとを纏めて一つのファイルとしている。

**【0008】** 発音時間計算部4は文章分解部1からの発音制御データに基づいて規則音声合成部2と全く同じアルゴリズムを用いて音声を合成する際の各音節が発音されるまでの時間をそれぞれ計算する。つまり、入力された文章に対してそれが規則音声合成部2で音声合成されて発音出力される際に、文章の先頭を起点にしてその文章を構成する各音節の切れ目のタイミングをそれぞれ推定してその結果を画像表示制御部5に出力する。

【0009】画像表示制御部5は発音時間計算部4からのタイミング信号に基づいて、各音節の発音タイミングが到来したときにその該当する音節の口形符号に対応する口形画像が画像メモリ6から選択されて出力されるよう画像表示制御を行う。すなわち、規則音声合成部2で発音される音声に対して画面に表示される話者の口の動きが一致するよう、つまり合成音声と顔動画像との同期がとれるように同期制御を行うものである。

【0010】パラメータ入力部7は規則音声合成部2で合成する音声の声質、顔動画像の画面上での表示場所、表示倍率等の各種パラメータをキーボード等を用いて入力する部分であり、合成音声に関するパラメータは規則音声合成部2に渡され、また顔動画像に関するパラメータは画像表示制御部5と画像メモリ6に渡される。

【0011】このように構成した装置の動作を説明する。文章情報が入力されると、文章分解部1でその文章情報が解析されて音素データがまとめて規則音声合成部2に渡されて合成音声により発音出力される。この発音動作に並行して、音素データが音声／口形変換部3で口形符号の系列に変換される。また発音時間計算部4では音素データから各音節の切れ目の時間が推定され、この時間データが画像表示制御部5に渡される。画像表示制御部5では各音節の発音タイミングに口形符号のタイミングを合わせて、画像メモリ6上に展開された各口形符号の画像のうちから音声／口形変換部3で求まった口形符号に対応した顔動画像データがVRAMに転送されるようにし、このVRAMを介して表示装置の画面上に話者の顔動画像を表示する。これにより文章情報は、それを実際に発音した合成音声とその合成音声に口の動きのタイミングがあった話者の顔動画像とによるメッセージとして受信者に伝えられることになる。

【0012】この図4の装置は、規則音声合成部2に従来からある小型の音声合成ユニットを利用し、それ以外の部分にはパソコン等を用いることにより、小型で経済的なシステムとして実現することができる。

### 【0013】

【発明が解決しようとする課題】かかる音声・顔動画像出力装置をパソコン上で実現させる場合、処理量削減のため、上述したように合成画像を予め作成しておいてそれらの画像を入力された文章に応じて切り換えて表示することが一般に行われている。これらの装置において合成音声と顔動画像を生成するにあたっては、声質、画面上での画像の表示場所、表示倍率などのパラメータは、表示するシステムに初期値として予め設定されたもの（パラメータ入力部7で予め入力されたもの）が使われる。

【0014】このように従来の装置では合成音声の声質と顔動画像の生成態様を受信側で予め設定しておくものであるが、それら予め登録されてある顔画像の人物と声

質が例えればメッセージに対して釣り合っていないような場合、それをみる人に不自然な感じを与えることになる。

【0015】また、この装置を電子メールなどに用いた場合などに代表されるように、文章情報と合成画像を作った人とその文章情報を実際に音声と動画像で表示して見る人が異なる場合、文章情報と合成画像を作った人が希望するような声質や画像の大きさで、受信側において発音・画像表示されるとは限らず、この結果、送り側の人の意図とは全く違う印象を受信側の人に入れてしまう可能性がある。

【0016】つまり従来の装置では、音声と動画像でメッセージを伝えるにあたっての声質や顔の容貌などから表示側の人が受ける印象は表示側で予め設定したパラメータによって決まってしまうことになり、情報の作成側の人が意図した印象表現を表示側の人に的確に伝えることができなかった。

【0017】本発明はかかる問題点に鑑みてなされたものであり、その目的とするところは、表示側で文章情報に基づいて合成音声あるいは顔合成画像を表示するにあたり、その文章等の作成側の人が意図した通りの表示が可能となるようにすることにある。

### 【0018】

【課題を解決するための手段】図1は本発明に係る原理説明図である。本発明においては、一つの形態として、任意の文章情報からそれに対応する合成音声および該合成音声に合わせて口が動く人物の顔の合成動画像を生成する合成画像表示システムにおいて、文章情報の作成側において顔の合成画像を作成する際に、表示側における合成音声と合成動画像の生成態様を決めるための各種パラメータを合成画像に付加して表示側に渡すように構成されたことを特徴とする合成画像表示システムが提供される。

【0019】上記の各種パラメータは合成音声の声質、合成動画像を表示する際の表示倍率、表示位置を含むパラメータとすることができる。

【0020】また本発明においては、他の形態として、受信した伝送データを合成動画像データ、文章情報、各種パラメータに分離する伝送データ入力手段と、文章情報に基づいて合成音声を生成し出力する音声合成手段と、伝送データ入力手段で分離された合成動画像データをファイリングする画像メモリと、文章情報をその文章情報を発声したときの一連の口形の動きを表す口形符号の系列に変換する変換手段と、文章情報に基づいて該音声合成手段から出力される合成音声の各音節の発音時間を計算して各音声の切れ目のタイミングを推定する発音時間計算手段と、発音時間計算手段で推定した各音節の切れ目のタイミングで表示画像を該変換手段からの口形符号に対応した口形画像に切り換える制御を行う表示制御手段と、伝送データ入力手段で分離された各種パラメータを合成画像に付加して表示側に渡すように構成されたことを特徴とする合成画像表示システムが提供される。

ータを対応する内部回路に送るパラメータ入力手段とを備えた合成画像表示システムが提供される。

#### 【0021】

【作用】本発明の合成画像表示方式においては、送信側において、表示に必要な顔画像を合成した際、その同じデータに表示側で合成した人が希望する合成画像の表示倍率や表示位置、合成音声の声質、その他のパラメータを埋め込む。表示側では、システムの初期値としてその画像データに埋め込まれた値を用いる。これにより、顔画像を合成した人の意図した通りに表示システム側で合成音声と合成画像を生成することができる。

【0022】また本発明の他の形態の合成画像表示システムにおいては、伝送データ入力手段で受信した伝送データを合成動画像データ、文章情報、各種パラメータに分離し、文章分解手段で文章情報を解析して発音制御データを生成し、音声合成手段でこの発音制御データに基づいて合成音声を生成し出し、受信した合成動画像データを画像メモリにファイリングし、変換手段で発音制御データを口形符号の系列に変換し、発音時間計算手段で発音制御データに基づいて音声合成手段で発音される各音節の発音時間をそれぞれ計算して各音節の切れ目のタイミングを推定し、画像表示制御手段で各音節のタイミング信号に合わせてその音節の口形画像を画像メモリから読み出すように制御し、受信した各種パラメータをパラメータ入力手段で対応する内部回路に送って合成音声と合成画像の生成様式を文章作成側の人が意図したものとなるようとする。

#### 【0023】

【実施例】以下、図面を参照して本発明の実施例を説明する。図2は本発明の一実施例としての合成画像表示システムによる音声・顔動画像出力装置が示される。図2において、文章分解部1、規則音声合成部2、音声／口形変換部3、発音時間計算部4、画像表示制御部5、画像メモリ6は前述の従来例で説明したものと同じものである。

【0024】従来装置との相違点として、送信側から送られてきた伝送データには、本来の文章情報の他に、送信側の人が受信側で合成され表示されることを希望する顔画像と口形画像等の合成画像データ、さらにその人が希望する画面上での表示倍率、表示位置、声質、その他のパラメータが合成画像データに埋め込まれている。

【0025】図3には送信側においてこれらのパラメータを伝送データに埋め込むための処理の概念が示される。受信側での表示を希望する顔画像の原画像に基づいて顔モデルへのマッピングを行い、各口形のパラメータを用いて合成画像データを作成し、これに表示倍率、表示位置、声質、その他の受け側に与える印象に係わるパラメータを埋め込む。これとは別に文章情報を作成し、双方を伝送データとして受信側に送る。この場合、パラメータが埋め込まれた合成画像データを一度送ってしま

えば、後は文章情報を繰り返し送るだけでよい。

【0026】受信側ではこの伝送データは伝送データ入力部8に入力され、ここで、文章情報、合成画像データ、各種パラメータに分離され、文章情報は文章分解部1に、合成画像データは画像メモリ6に、各種パラメータはパラメータ入力部7にそれぞれ送られる。

【0027】パラメータ入力部7はこの各種パラメータを受け取ると、この各種パラメータを調べて、音声合成に関する声質等のパラメータは規則音声合成部2に、画像の表示倍率、表示位置等の画像に関するパラメータは画像表示制御部5と画像メモリ6にそれぞれ送る。

【0028】このように構成することで、受信側では、表示すべき顔画像とシステムの初期値として埋め込むパラメータとして、送信側の人が希望した顔画像と、表示倍率、表示位置、声質、その他のパラメータを用いることができる。よって送信側の人の意図通りの音声と画像で受信側の表示システムにメッセージを表示させることができる。

【0029】本発明の実施にあたっては種々の変形形態が可能である。例えば、上述の実施例では口形として7種類の画像を用いる場合について説明したが、もちろん本発明はこれに限られるものではなく、より自然に近い口の動きを合成するためにはこの口形の画像の種類をさらに増やしてもよい。また上述の実施例では表示側で合成する顔画像の動き部分として口領域の動きを取り上げたが、これに限られるものではなく、例えば口の動きに加えて、文章に合わせて目の動きなども変化させるようすれば、より表情豊かなAVメッセージを受け側に送ることができる。

【0030】また上述の実施例では本発明をAV電子メールに適用した場合について説明したが、本発明はこれに限られるものではなく、音声・顔動画像出力装置単体に適用することも可能であるし、あるいは、例えば音声認識技術によりリアルタイムに発聲音声の音素の認識が可能となれば、通常の電話をかけるだけで受信者側に話し手の顔の表情も動画像で表示できるという擬似テレビ電話等のサービスに適用することも可能である。

#### 【0031】

【発明の効果】以上に説明したように、本発明によれば、受信側で文章情報に基づいて合成音声あるいは顔合成画像を表示するにあたり、その文章の送り側の人が意図した通りの表示が可能となる。

#### 【図面の簡単な説明】

【図1】本発明に係る原理説明図である。

【図2】本発明の一実施例としての合成画像表示システムによる音声・動画像出力装置を示す図である。

【図3】実施例システムによる送り側での処理概念を説明する図である。

【図4】従来の音声・動画像出力装置を示す図である。

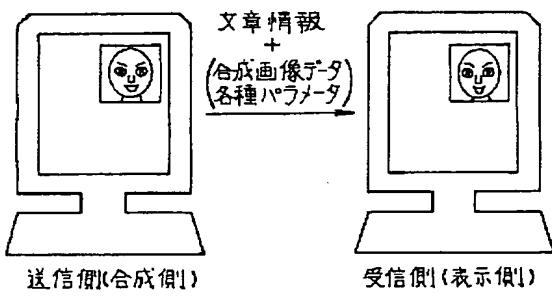
【符号の説明】

- 1 文章分解部
- 2 規則音声合成部
- 3 音声／口形変換部
- 4 発音時間計算部

- 5 画像表示制御部
- 6 画像メモリ
- 7 パラメータ入力部
- 8 伝送データ入力部

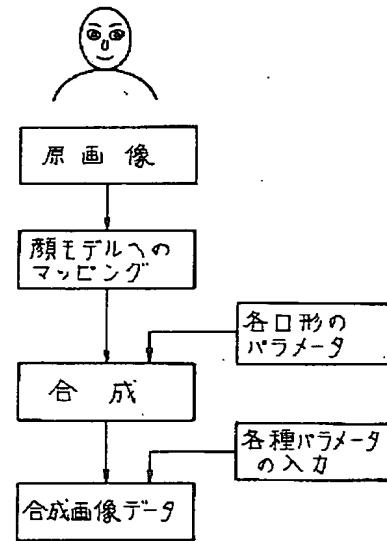
【図1】

## 本発明に係る原理説明図



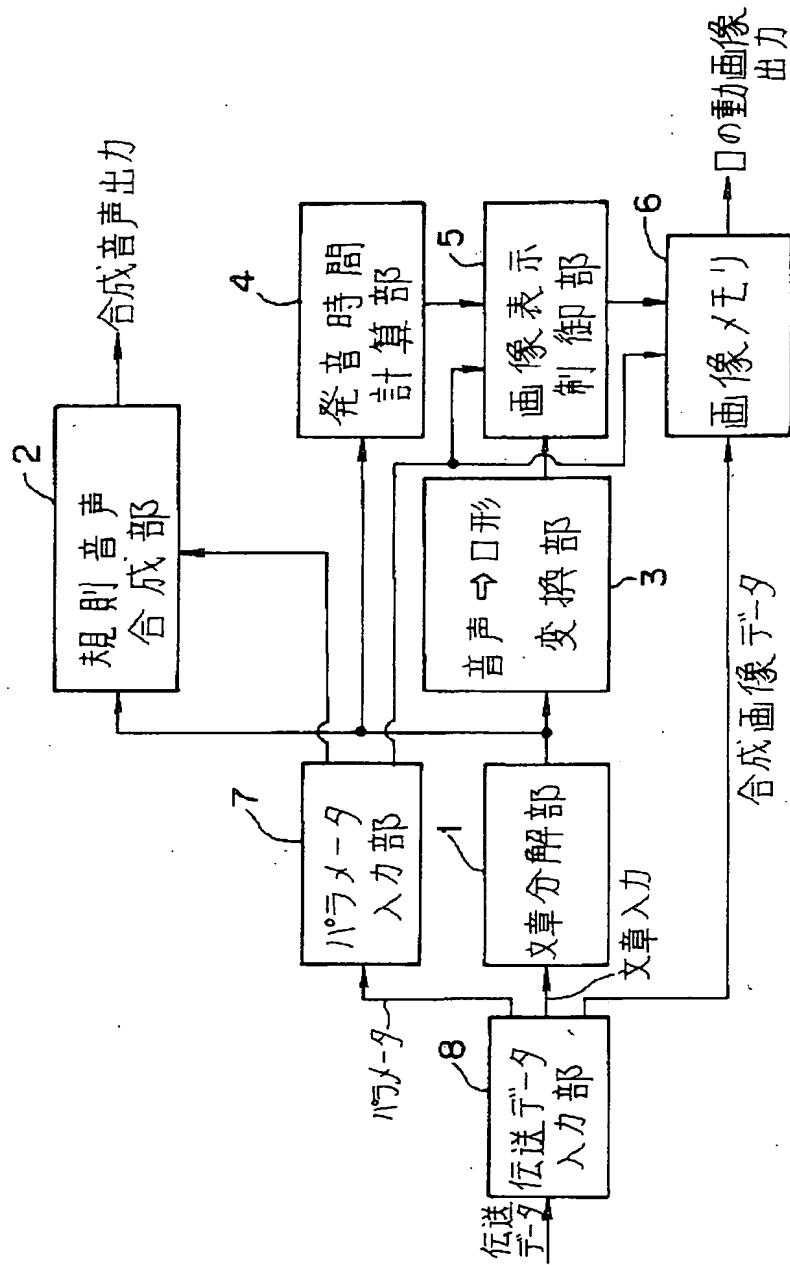
【図3】

## 送信側の処理



【図2】

## 本発明の実施例



【図4】

往來例

